

木構造を用いた音韻連鎖統計モデル*

7 N-4

田本 真詞 伊藤 克亘 田中 穂積†
(東京工業大学 工学部)‡

1はじめに

音声認識処理中の音韻認識率は次第に向かっているが、100%の認識率を持つような音韻モデルはない。そこで認識の性能を向上させるための手段として、様々な言語情報を利用することが考えられる。これらの情報のひとつとして、音韻連鎖に関する統計情報が注目されている。実際、音韻連鎖の統計的言語モデルが音声認識に有効に働くことが知られている[1][2]。本研究では、木構造を用いて、情報量が最大になるように、文脈に応じて参照する音韻連鎖の長さを動的に変化させる、精度が良く信頼性の高い統計モデルを提案する。さらに、この木構造モデルとN-gramの情報量を比較する実験を行ない、本モデルの有効性を検証する。

2 音韻連鎖の統計モデル

ある音韻列 $P = p_1, p_2, \dots, p_n$ が生成される確率は、N-gram を用いて近似的に次のように表される。

$$P(P) = \prod_{i=1}^n P(p_i | p_{i-N+1}, p_{i-N+2}, \dots, p_{i-1}) \quad (1)$$

この近似方法では、Nが大きくなるにつれ利用できる文脈情報が増大し、推定精度が向上すると考えられる。ところがN-gramの確率は、データ中の出現頻度から算出するので、Nを大きくすると $p_{i-N+1}, p_{i-N+2}, \dots, p_{i-1}$ の組合せがべき乗で増加する。このため、有限量のデータから生成確率を求めた場合、統計モデルとしての信頼性が損なわれ、かえって推定精度を低下させることになる。このような問題に対して、削除補間法によるN-gramの線形結合[3]や、出現頻度の小さいN-gramのback-offスマージング法による推定[4]、等の様々なデータの補間法がある。しかし、このような補間を行なっても3-gramを大きく上回る成果は見い出されていない[5]。この他に、決定木による木構造言語モデルを使い、同等の3-gramよりパープレキシティーを改善した報告[6]がある。

3 木構造を用いた音韻連鎖の統計モデル化

文脈依存性の高い、すなわち音韻連鎖の長い統計モデルでは、特定の文脈における音韻の生成確率をより正確にとらえることができるため、モデルの精度は高い。しかし、音韻連鎖の組合せが多くなり、信頼性の高いモデルを生成するためには多量のデータが必要になる。逆に文脈依存性の低い、音韻連鎖の短いモデルでは、生成確率が様々な文脈の混合となるために統計モデルの正確さに欠けるが、小量のデータで信頼性の高いモデル化ができる。

そこで音韻連鎖の生成確率をモデル化するためには、ある音韻連鎖とそのデータの量を考慮する必要がある。つまり出現頻度の大きい音韻連鎖をより長くすれば、全体の信頼性を損なうことなく、精度の高いモデル化ができる[7]。

日本語の音韻的な特性から、文中に頻出する付属語や漢字など語構成要素[8]のレベルで長さ3ないし4の音韻連鎖がある程度決定されていると考えられる。このような連鎖は、局所的な領域に分布するので、ある音韻の生成確率を求めるには、その直近の音韻連鎖がわかればよい。

このように条件に応じて特定の音韻連鎖を伸長させるモデルには、連鎖の長さを順次変化させることのできる木構造が適している。また音韻連鎖の生成確率を計算する場合も、直前の音

韻連鎖の生成確率がわかっているれば、生成確率の再計算を大幅に削減することができる。

例えば、3-gram ($P(p_x | p_1, p_2)$) を木構造で表現すると、リーフからルート方向に音韻 p_1, p_2 に相当するノードが配置される。このとき、リーフ側の p_1 には、 $P(p_x | p_1, p_2)$ の3-gramの各生成確率が、ルート側の p_2 には、 $P(p_x | p_2)$ の2-gramの各生成確率が、ルートノードには、 $P(p_x)$ の1-gramの生成確率が蓄積される。

このような構造を用いれば、認識時に木をたどることによって、目的の音韻連鎖の生成確率を得ることができる。また、 p_1 の先に順次リーフを追加することにより、音韻連鎖を先行する方向に伸長することができる。

そこで木構造を用いれば、精度の低いモデルから順次精度の高いモデルへと成長させることができる[9]。

3.1 木構造の生成アルゴリズム

統計モデルの信頼性の高さを保ちながら精度を向上させるために、木構造の持つ情報量を最大化するようにリーフの追加を行なう。これは、木に含まれるすべてのノードについて新たなノードを追加した後の情報量の増分を求め、これを最大にするノードを探索することに等しい。

あるノードに新たなリーフを追加する手続きは、次のようになる。

- 生成確率の条件部の音韻連鎖が l_i に相当するようなリーフ l_i を考える。ルートノードならば l_i の長さは 0 である。このノードの持つ情報量を H_{l_i} とする。

$$H_{l_i} = \sum_j -P(l_i \cdot p_j) \cdot \log P(l_i \cdot p_j) \quad (2)$$

- リーフを追加したときの情報量の差分を計算する。

- 新たに追加されたリーフに相当する音韻を p_k としたとき、 p_k の追加によって得られる情報量 $H_{p_k \cdot l_i}$

$$H_{k \cdot l_i} = \sum_j -P(p_k \cdot l_i \cdot p_j) \cdot \log P(p_k \cdot l_i \cdot p_j) \quad (3)$$

- リーフの追加により、 l_i の持つ情報量が変化する。変化後の情報量 H'_{l_i}

$$H'_{l_i} = \sum_j -P(p_{x \neq k} \cdot l_i \cdot p_j) \cdot \log P(p_{x \neq k} \cdot l_i \cdot p_j) \quad (4)$$

分割後の情報量の増分は $H'_{l_i} + H_{k \cdot l_i} - H_{l_i}$ で求められる。

- すべての音韻 p_k ($k = 1 \dots n$) について情報量の差分を求める。同様にすべてのリーフノード l_i ($i = 1 \dots N$) に対して情報量の差分を求めておく。
- すべてのノードについて、すべての分割に関する情報量の差分を比較し、差分が最小となるノードを新たに追加する。
- 新しいリーフノードを l_{i+1} とし、1. にもどる。

3.2 音韻連鎖の生成確率

このようにして生成された統計モデルにおける音韻連鎖の生成確率は、次のように表される。

*A Tree-based Stochastic Phone Sequence Modeling.

†Masafumi Tamoto, Katunobu Itou and Hozumi Tanaka

‡Tokyo Institute of Technology

リーフノードにおける生成確率 リーフノードを l_i とするととき、音韻連鎖 l_i に続いて音韻 p_j が生成される確率 $P_{l_i}(j)$

$$P_{l_i}(j) = P(p_j | l_i) \quad (5)$$

リーフ以外のノードにおける生成確率 リーフ以外のノードにおける生成確率は、音韻が後方に伸長している分だけ変化する。この値は、条件つき確率だけでは求められない。ノード n_i において先行する音韻の集合が $p_{k \in A}$ であるとき、音韻連鎖 $p_{k \in A} \cdot n_i$ の生成確率は、学習データにおける音韻連鎖 P の出現回数を $C(P)$ とすることで次のように表せる。

$$P_{n_i}(j) = \frac{C(n_i \cdot p_j) - C(p_{k \in A} \cdot n_i \cdot p_j)}{C(n_i) - C(p_{k \in A} \cdot n_i)} \quad (6)$$

4 統計モデルの評価

木構造を生成するのに用いた訓練セットとは別の評価用セットを使い、1文あたりの平均生成確率、テストセットバープレキシティー、及び木構造の持つ情報量を N-gram (N = 1...5) と比較する。なお、テストセットバープレキシティは、次の式で表される。

$$F_T(p) = \left(\prod_{i=1}^n \frac{1}{P(p_i | p_{i-N+1}, p_{i-N+2}, \dots, p_{i-1})} \right)^{\frac{1}{n+1}} \quad (7)$$

4.1 N-gram との比較

表 1: テストセットの coverage

gram 長	1-gram	2-gram	3-gram	4-gram	5-gram
coverage	100%	100%	99.9%	98.9%	95.0%

表 2: 木構造モデルの音韻連鎖数、平均長と coverage

リーフ 数	1000	2000	5000	10000	20000
平均 長	1.88	2.22	2.62	2.89	3.20
coverage	100%	100%	100%	100%	100%

表 3: 一文あたりの平均生成確率

N-gram					
gram 長	1-gram	2-gram	3-gram	4-gram	5-gram
生成確率	5.13E-2	1.09E-1	1.40E-1	1.66E-1	1.53E-1

木構造モデル					
リーフ 数	1000	2000	5000	10000	20000
生成確率	7.80E-2	9.86E-2	1.23E-1	1.35E-1	1.51E-1

表 4: 一文あたりのバープレキシティ

N-gram					
gram 長	1-gram	2-gram	3-gram	4-gram	5-gram
	19.7	9.28	7.35	6.75	12.1

木構造モデル					
リーフ 数	1000	2000	5000	10000	20000
	13.0	10.3	8.33	7.56	6.79

表 5: 統計モデルの持つ情報量

N-gram					
gram 長	1-gram	2-gram	3-gram	4-gram	5-gram
		3.15	2.74	2.29	1.83

木構造モデル					
リーフ 数	1000	2000	5000	10000	20000
	8.71	10.6	13.4	15.9	18.8

4.2 音韻連鎖長の制限

リーフノードが増加し、音韻連鎖の平均長が大きくなるにつれ、木構造モデルの持つ情報量は増加していくが、最初に述

べたようにモデル全体の信頼性は低下することがある。これを防ぐために、木構造の成長がある評価をもとに制限しなければならない。このために、あるリーフ l_n における音韻連鎖が p_1, p_2, \dots, p_n であるとき、生成確率が最大になるように音韻連鎖の長さを制限する。すなわち、ある音韻連鎖の生成確率をつぎのように表す。

$$P'_{l_n}(j) = \underset{1 \leq i \leq n}{\operatorname{argmax}} P(p_j | p_i, \dots, p_n) \quad (8)$$

5 結論

リーフ数 20,000 ではなく 5-gram 程度の平均生成確率、及びテストセットバープレキシティーを得ることができた。ただし、テストセットに対する coverage は、木構造モデルの方が優れている。実際の音声認識に用いるときは、coverage の低いモデルは信頼性が低下することが知られており、その点から考えると本手法は、音声認識システムで使用した場合、良好な結果を得ることができるものと期待できる。

今後、実際の音声認識システムを用いてモデルの有効性を検証したい。

参考文献

- [1] 川端、花沢、伊藤、鹿野、「HMM 音韻認識における音節連鎖統計情報の利用」、信学技法、SP-89-110 (1990)
- [2] 村上仁一、荒木哲郎、池原悟、「2 重マルコフ連鎖確率モデルを使用した単音節音声入力」、電子情報通信学会技術報告 SP88-29 (1988)
- [3] F.Jelinek, R.Mercer, "Interpolated Estimation of Markov Source Parameters from Sparse Data", *Pattern Recognition in Practices*, E.S.Gelsema and L.N.Kanal, ed., North-Holland Publishing Company (1980)
- [4] S. Katz, "Estimation of Probabilities from Sparse Data for the Language Model Component od a Speech Recognizer", IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, ASSP-34(3) 1989
- [5] F.Jelinek, "UP FROM TRIGRAMS!", Proc.EuroSpeech 91 (1991)
- [6] Lalit Bahl, Peter F. Brown, Peter V. Souza, "A Tree-Based Statistical Language Model for Natural Langae Speech Recognition", IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. 37, NO.7, JULY 1989
- [7] 伊藤克亘、「日本語の統計的な振舞いを利用した連続音声認識」、修士論文、東京工業大学、(1990)
- [8] 水谷静夫、田嶋一夫、佐竹秀雄、野村雅昭、石井雅彦、樺島忠夫、「文字・表記と語構成」、朝倉日本語新講座、(1987)
- [9] 速水悟、田中和世、「木構造音韻モデルによる未知音素文脈中の音響的変動の予測と評価」、電子通信情報学会、SP90-64 (1990)